Reproducing Sound Fields Using MIMO Acoustic Channel Inversion

Mihailo Kolundžija, Christof Faller, and Martin Vetterli

***Abstract*—Sound fields are essentially band-limited phenomena, both temporally and spatially. This implies that a spatially sampled sound field respecting the Nyquist criterion is effectively equivalent to its continuous original. We describe Sound Field Reconstruction (SFR)—a technique that uses the previously stated observation to express the reproduction of a continuous sound field as an inversion of the discrete acoustic channel from a loudspeaker array to a grid of control points. The acoustic channel is inverted using truncated singular value decomposition (SVD) in order to provide optimal sound field reproduction subject to a limited effort constraint. Additionally, a detailed procedure for obtaining loudspeaker driving signals that involves selection of active loudspeakers, coverage of the listening area with control points, and frequency-domain FIR filter design is described. Extensive simulations comparing SFR with Wave Field Synthesis show that on average, SFR provides higher sound field reproduction accuracy.**

***Index Terms*—Sound field reproduction, Wave Field Synthesis, MIMO equalization, Loudspeaker arrays**

I. Introduction

The first spatial sound reproduction systems date back to the work of Blumlein [1] on stereo systems in the first half of the last century. The successful two-channel stereo principle—still used widely today—was extended to the four-channel quadraphonic system [2] with the aim of providing full-circle spatial reproduction, but it was quickly abandoned due to unsatisfactory front localization, technical issues, and format incompatibilities. Surround systems using a higher number of channels, such as 5.1 and 7.1, are based on the observation that accurate localization of the sound coming from the front is more important, and they use more loudspeakers in the front of the listener for improved frontal localization. Additionally, loudspeakers on the side and behind the listening position are used for providing ambience and side/rear localization.

All previously mentioned surround sound systems create sound fields with correct spatial attributes only within a narrow listening area called the "sweet spot". The problem of extending the listening area was addressed by two notable surround sound systems: Ambisonics (e.g., see [3], [4]) and Wave Field Synthesis (e.g., see [5]–[7]). Both approaches are based on an attempt to reproduce a *desired sound field* in an extended listening area.

The theoretical foundations of Ambisonics were laid down in the '70s (e.g., see [3], [4], [8]), primarily for circular and spherical loudspeaker arrangements. At the heart of the ambisonic reproduction technique is the sound field mode matching in one, central listening spot. In this particular case, mode matching implies matching orthogonal components, such as cylindrical or spherical harmonics, of desired and reproduced sound fields. The early ambisonic systems suffered from a limited sweet-spot size, particularly at medium and high frequencies [9], due to the use of modes of low order, an insufficient number of loudspeakers, and far-field loudspeaker models. Later, Daniel *et al.* [10] provided extensions to the initial works on Ambisonics, considering higher order modes and modeling loudspeakers as point sources to more accurately account for propagation effects. Near-field higher-order Ambisonics was shown to have comparable performance to WFS for enclosing loudspeaker configurations [10], but for practical systems, it lacks recording support in the form of a wideband high-order sound field microphone.

Wave Field Synthesis (WFS) systems, on the other hand, are based on the Helmholtz integral equation (HIE). Roughly speaking, HIE shows how a desired sound field in a closed source-free (listening) domain can be reproduced by a continuous distribution of secondary monopole and dipole sources on the domain boundary. In the initial works of Berkhout *et al.* [5], [7], WFS was derived starting from a particular form of HIE—Rayleigh I and II integrals, where the listening domain is a half-space and secondary sources—monopole in the former and dipole in the latter—are distributed on the bounding plane. Since the reproduction in the horizontal plane is far more important than in the vertical plane [11], the creators of WFS focused on linear loudspeaker setups and to approximate the performance of planar source distributions. Using stationary phase approximation, they

The authors are with the School of Computer and Communications Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland (e-mail: {mihailo.kolundzija, christof.faller, martin.vetterli}@epfl.ch).

were able to derive the so-called $2^{1/2}$-dimensional WFS, which is able to approximately reproduce a desired sound field in the listening plane, while reproducing it exactly on a reference listening line. The initial WFS concept was extended by Start [12] to include curved loudspeaker distributions, and Verheijen [13] and de Vries [14] to reproduce arbitrarily directive sources and use directional loudspeakers, respectively.

More recently, Ahrens and Spors proposed in [15] an approach called Spectral Division Method (SDM), which formulates the sound field reproduction using planar and linear loudspeaker distributions as a spatio-temporal spectral inversion. They have shown that in some particular cases, such as the reproduction of plane waves using monopole sources, one is able to obtain a correct closed-form solution for loudspeaker driving signals.

However powerful as theoretical tools, the mentioned approaches for sound field reproduction need to cope with limitations imposed by systems used in practice. Namely, they need to be applied to discrete loudspeaker distributions of limited spatial support, varying directivity and possibly multi-path propagation, while listening domains are of finite size. Some of these issues have been addressed by Verheijen [13], who proposed the use of geometry-based loudspeaker subset selection and spatial tapering towards the loudspeaker array edges, to mitigate the impairments due to spatial aliasing and diffraction, respectively. Corteel [16], [17] used a similar loudspeaker selection method, while Spors [18] proposed a slightly different approach based on the direction of sound intensity vectors.

More recent approaches for computing loudspeaker filters for sound field reproduction use a discretization of the listening area and numerically solve a discrete optimization problem. This is partly due to unsatisfactory performance of analytical solutions in practical problems, and partly to avoid restricting the reproduction setup (e.g., having calibrated loudspeakers of prescribed directivity, free-field conditions etc.). One of the earliest numerical approaches, by Kirkeby and Nelson [19], addresses reproduction of plane waves based on pseudo-inversion of a multichannel acoustic propagation matrix. Similar to [19], but applied to WFS for the purpose of room equalization and directive source reproduction, are the works of Corteel [16], [17], where the control points are arranged on four listening lines. Gauthier and Berry [20] use only four control points arranged as a quadrupole to compute loudspeaker filters that optimize a cost function consisting of the reproduction error and deviation from the initial WFS driving signals.

In this paper, we present an approach denoted as *Sound Field Reconstruction* (SFR) [21], [22]. In essence, SFR is designed to optimally reproduce a desired sound field in a *given listening area* for a *given finite setup*, while keeping loudspeaker driving signals well behaved in order to respect physical constraints. SFR has three important aspects:

- Design of a control point grid covering the listening area
- Selection of active loudspeakers and a subset of control points based on position of the reproduced source and the geometry of the reproduction setup
- Computation of loudspeaker filters using MIMO channel inversion.

The grid of control points inside the listening area used in SFR is designed following the equivalence of sound reproduction in a continuous and sampled spatial domain, proved in the next section. Furthermore, SFR uses only those loudspeakers that mostly contribute to the sound field reproduction, as this strategy is already known to mitigate high-frequency spatial aliasing problems characteristic of loudspeaker arrays [13], [16], [23]. The active loudspeakers are selected based on geometry, similar to [13]. Also, in order to avoid over-fitting, the control points where desired sound field evolution cannot be locally matched are discarded following similar geometrical considerations. Finally, SFR uses a variant of MIMO channel pseudo-inversion with truncated singular value decomposition (SVD). This technique allows graceful degradation of sound field reproduction performance when the MIMO channel matrix is ill-conditioned, while keeping loudspeaker filters within practical physical system constraints. Being both setup and listening area optimized, SFR is able to achieve higher sound field reproduction accuracy than WFS, as will be shown with simulations.

The paper is organized as follows. Section II describes theoretical and practical aspects of SFR. Section III presents extensive evaluation of SFR and its comparison with WFS. Practical considerations for realizing SFR systems are discussed in Section IV. Conclusions are presented in Section V.

II. Sound Field Reconstruction

As mentioned in the introduction, Sound Field Reconstruction (SFR) is a spatial sound reproduction approach which is based on the spectral properties of the plenacoustic function shown by Ajdler *et al.* [24]. More particularly, it is based on the essential spatial band-limitedness[1] of the sound field that emanates from

[1]*Essentially* band-limited function in this context refers to a function which has most of its energy in a finitely-supported spectral region, while the energy outside of that region decays exponentially.

temporally band-limited sound sources. This section provides a description of sampling and interpolation of the plenacoustic function, and shows how these can be used for sound field reproduction with arbitrary reproduction setups. It also describes practical extensions that help to improve the sound field reproduction with SFR in specific finite domains, and briefly presents the design of discrete-time loudspeaker filters for SFR.

A. Plenacoustic sampling and interpolation

In the most general sense, the plenacoustic function $p(\boldsymbol{r}, t)$ describes a sound field in space and time, irrespective of the sources evoking it. In a particular case where the sound field is evoked by a point source located at $\boldsymbol{r}'$ emitting a Dirac pulse, the plenacoustic function equals the Green's function $h_{\mathbf{r}'}(\mathbf{r}, t)$, i.e., the spatio-temporal impulse response of the acoustical medium from point $\mathbf{r}'$ to point $\mathbf{r}$.

The changes of the plenacoustic function in space at a temporal frequency ω can not happen at an arbitrary rate, but are limited by ω according to the relation [24]

$$k^2 \leq \tfrac{\omega^2}{c^2}\,, \tag{1}$$

where k is the spatial frequency and c is the speed of sound propagation.[2]

Based on the observation (1), one can define a minimum spatial sampling frequency for a sound field of limited temporal bandwidth. If the maximum temporal frequency of the sources evoking the sound field is equal to ω_m, then a spatial sampling frequency of $k_s = 2\omega_m/c$ is sufficient for representing the sound field. This observation extrapolates to a large extent to finite spatial segments (e.g., finite-length lines and finite-area rectangles), as shown in [24].

The possibility of sampling a sound field has an implication that is useful in the context of SFR. Namely, it suggests that correct reproduction of a sound field on a grid of points can guarantee correctness of reproduction between the grid points. Without loss of generality, we show this result for the xy plane in Theorem 1. However, we first give a lemma which simplifies proving said theorem.

Lemma 1. *If two functions $f(x, y, t)$ and $h(x, y, t)$, both band-limited with the maximum temporal frequency ω_m and maximum spatial frequency $k_m = \omega_m/c$ ($k^2 = k_x^2 + k_y^2 \leq k_m^2$), are identical on a 2D grid*

$$(n\Delta x,\ m\Delta y)\,, \quad n, m \in \mathbb{Z}\,,$$

where

$$\Delta x \leq \tfrac{k_m}{\pi},\ \Delta y \leq \tfrac{k_m}{\pi}\,,$$

then they are identical everywhere.

Proof: The proof follows from the fact that a band-limited function (e.g., a sound field) is uniquely defined by its samples on a grid satisfying the Nyquist criterion. Since the functions $f(\mathbf{r}, t)$ and $h(\mathbf{r}, t)$ are both band-limited with the same spectral support, and have identical values on a sampling grid satisfying the Nyquist criterion, they must be identical everywhere. ■

Theorem 1. *Consider two sets of temporally band-limited sources, $S_1 = \{s_{11}(t), ..., s_{1k}(t)\}$ and $S_2 = \{s_{21}(t), ..., s_{2l}(t)\}$, each with the maximum temporal frequency ω_m. Let each source $s_{ij}(t)$ be placed at a location $\boldsymbol{r}_{ij}$. Assume that the spectral support of each Green's function $h_{\boldsymbol{r}_{ij}}(\boldsymbol{r}, t)$, evaluated in the xy plane, is confined to the double cone*

$$k^2 = k_x^2 + k_y^2 \leq \tfrac{\omega^2}{c^2}\,.$$

For each separate source $s_{ij}(t)$, the sound field evoked by it in the xy plane is given by

$$p_{ij}(x, y, t) = \int h_{\boldsymbol{r}_{ij}}(x, y, \tau) s_{ij}(t - \tau) \mathrm{d}\tau\,.$$

Further, let $P_1(x, y, t)$ and $P_2(x, y, t)$ be the superposed sound fields of the sources in S_1 and S_2, respectively, given by

$$P_i(x, y, t) = \sum_j p_{ij}(x, y, t)\,.$$

The two sound fields, $P_1(x, y, t)$ and $P_2(x, y, t)$, are identical in the entire xy plane if they are identical on a grid given by

$$(n\Delta x,\ m\Delta y)\,, \quad n, m \in \mathbb{Z}\,, \tag{2}$$

with

$$\Delta x \leq \tfrac{\omega_m}{c\pi},\ \Delta y \leq \tfrac{\omega_m}{c\pi}\,.$$

Proof: The sound field of each source is band-limited in time and space, with maximum temporal and spatial frequencies ω_m and $k_m = \omega_m/c$, respectively. Consequently, the sound fields $P_1(x, y, t)$ and $P_2(x, y, t)$, being superpositions of functions band-limited in space and time, are also band-limited with the same maximum frequencies. Their equality follows from Lemma 1. ■

Note also that even if the spatio-temporal spectrum of Green's function is not confined to the double cone defined by $k^2 \leq \omega^2/c^2$, its propagating part is.[3] Therefore,

[2] The given limit on the spatial variations of sound pressure is not entirely correct, but it is to a large extent when sources are not inside of or very close to the considered spatial domain.

[3] As mentioned previously, the propagating part contains essentially the entire energy of a sound field.

it follows that the propagating parts of two sound fields are equal if they are equal on the grid given by (2).

Based on this observation, the problem of reproducing the sound field that emanates from temporally band-limited sources with maximum temporal frequency ω_m is equivalent to reproducing the sound field on a grid of control points spaced at or above the Nyquist spatial sampling frequency $k_s = 2\omega_m/c$. In the case of practical sound field reproduction with an array of loudspeakers, the listening area, and thus also the control grid, are finite. Consequently, sound field reproduction can be expressed as a multiple-input multiple-output (MIMO) problem.

B. Sound Field Reconstruction using MIMO channel inversion

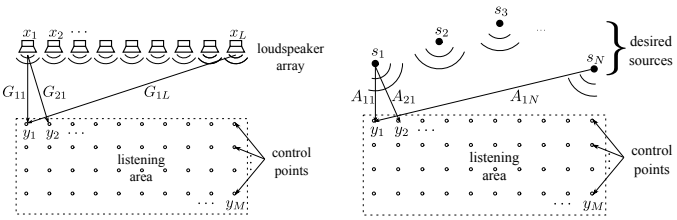
MIMO channel inversion is a standard problem that reappears in many multichannel sound applications, such as multi-point room equalization, sound field reproduction, and beamforming (e.g., see [16], [19], [25]). For the sake of completeness, we will present the MIMO channel inversion problem and the particular solution used in SFR.

The problem of MIMO channel inversion in the context of sound field reproduction is illustrated in Fig. 1. The reproduction setup includes an array of L loudspeakers and a grid of M control points covering the listening area, illustrated in Fig. 1(a). In addition, as shown in Fig. 1(b), there is a desired acoustic scene that contains N sound sources that would evoke the desired sound field in the listening area.

Positions of loudspeakers, control points, and desired sources are known. The transfer function $A_{ij}(\omega)$ denotes the sound propagation channel between the jth desired source and ith control point. Similarly, $G_{ik}(\omega)$ denotes the sound propagation channel between the kth loudspeaker and the ith control point. Both $A_{ij}(\omega)$ and $G_{ik}(\omega)$ are known for all pairs desired source-control point and loudspeaker-control point, respectively, either through a theoretical model or through measurement.

The goal of the MIMO channel inversion in the context of SFR is the reproduction of the desired sound scene in M control points, i.e., computation of the loudspeaker driving signals that evoke the same signals at the control points as the original sound scene.

Note that the problem of multichannel inversion can be represented as a superposition of N independent sub-problems, each involving a single desired source. The loudspeaker signals can then be obtained by summing the contributions for each single-source sub-problem. Thus, without loss of generality, the following MIMO channel



(a) Reproduction setup (b) Desired sound scene

Fig. 1. Multichannel inversion problem overview.

inversion analysis is presented only for the first desired source.

Denote by $S_1(\omega)$, $X_j(\omega)$, and $Y_k(\omega)$ the Fourier transforms of signals of the desired source, the output of the jth loudspeaker, and the sound pressure at the kth control point, respectively. Furthermore, denote by $D_l(\omega)$ the signal at the lth control point in the desired sound scene containing only the first desired source.

The signals $D_i(\omega)$ are determined by the effects of the sound propagation paths from the desired source to the control points, and are described by the following product:

$$\mathbf{D}(\omega) = \mathbf{A}(\omega) S_1(\omega)\,, \tag{3}$$

where

$$\mathbf{D}(\omega) = [D_i(\omega)]_{M\times 1}$$
$$\mathbf{A}(\omega) = [A_{i1}(\omega)]_{M\times 1}\,.$$

On the other hand, the signals produced by the loudspeakers at the control points are determined by the sound propagation effects on the loudspeaker signals, and are given by

$$\mathbf{Y}(\omega) = \mathbf{G}(\omega)\mathbf{X}(\omega)\,, \tag{4}$$

where

$$\mathbf{Y}(\omega) = [Y_i(\omega)]_{M\times 1}$$
$$\mathbf{G}(\omega) = [G_{ij}(\omega)]_{M\times L}$$
$$\mathbf{X}(\omega) = [X_i(\omega)]_{L\times 1}\,.$$

The task of the multichannel inversion is to compute the signals $X_j(\omega)$ using the desired signal $S_1(\omega)$, i.e.,

$$\mathbf{X}(\omega) = \mathbf{H}_1(\omega) S_1(\omega)\,, \tag{5}$$

where

$$\mathbf{H}_1(\omega) = [H_{i1}(\omega)]_{L\times 1}\,,$$

such that the difference (error) between the vector $\mathbf{Y}(\omega)$ and vector $\mathbf{D}(\omega)$, corrected by a constant delay Δ

accounting for the propagation time differences or the modeling delay, is minimized. The multichannel inversion problem is illustrated in Fig. 2.

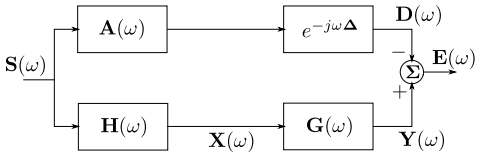


Fig. 2. Block diagram illustrating the MIMO channel inversion problem.

The solution which minimizes the error power, i.e., the mean squared error (MSE) solution, is given by [26]

$$\boldsymbol{H}_1(\omega) = e^{-j\omega\Delta}\boldsymbol{G}^{+}(\omega)\mathbf{A}_1(\omega)\,. \tag{6}$$

Since it uses a pseudo-inverse $\boldsymbol{G}^{+}(\omega)$ of the transfer matrix $\boldsymbol{G}(\omega)$, finding the pseudo-inverse of the matrix $\boldsymbol{G}(\omega)$ becomes the central problem of MIMO channel inversion.

The classical full-rank pseudo-inverse expression is given by

$$\mathbf{G}^{+}(\omega) = \left(\mathbf{G}^{H}(\omega)\mathbf{G}(\omega)\right)^{-1}\mathbf{G}^{H}(\omega)\,, \tag{7}$$

where the matrix $\boldsymbol{G}^{H}(\omega)$ is the conjugate-transpose of the matrix $\boldsymbol{G}(\omega)$. At low frequencies, where the condition number of the matrix $\boldsymbol{G}(\omega)$ is large (making it effectively low-rank), (7) gives filters with gains beyond the physical limitations of practical loudspeakers. The regularized pseudo-inversion used in [16], [25] is also of limited use, as it does not allow easy control of the trade-off between the reproduction accuracy and maximum filter gains.

Like a number of MIMO inversion solutions in acoustics (e.g., see [27]), we use a pseudo-inversion method based on the truncated singular value decomposition (SVD), which prunes singular values that are below a defined threshold ϵ (e.g., see [28]). In particular, if

$$\mathbf{G}(\omega) = \mathbf{U}(\omega)\mathbf{\Sigma}(\omega)\mathbf{V}^{H}(\omega) \tag{8}$$

is the SVD of the matrix $\mathbf{G}(\omega)$, then the pseudo-inverse of the matrix $\mathbf{G}(\omega)$ is given by

$$\mathbf{G}^{+}(\omega) = \mathbf{V}(\omega)\mathbf{\Sigma}^{+}(\omega)\mathbf{U}^{H}(\omega)\,, \tag{9}$$

where the matrix $\mathbf{\Sigma}^{+}(\omega)$ is obtained from $\mathbf{\Sigma}(\omega)$ by first setting to zero the singular values whose absolute values are below a defined threshold ϵ, replacing the other singular values by their reciprocal, and taking the matrix transpose in the end [28]. The threshold ϵ can be adapted to the matrix $\boldsymbol{G}(\omega)$, i.e., it can be set to a fraction of the largest singular value of $\boldsymbol{G}(\omega)$.[4]

[4]In the simulations presented in Section III, the threshold ϵ was 20 dB below the largest singular value of $\boldsymbol{G}(\omega)$.

At high frequencies, where all singular values of the matrix $\boldsymbol{G}(\omega)$ are larger than the threshold, this procedure gives the result identical to (7). However, at low frequencies, it gives near-optimal solutions while keeping the loudspeaker filter gains within practical limits. For a more detailed treatment of this MIMO channel inversion problem, see [21].

C. Practical extensions of Sound Field Reconstruction

1) Filter correction through power normalization on the reference line: All sound field reproduction approaches give loudspeaker driving signals that do not provide correct sound field reproduction accuracy above a certain aliasing frequency.

Although coming from the same physical limitations as in approaches such as WFS, and which are inherent to the geometry of the used loudspeaker array and location of the reproduced source, the high-frequency problems of SFR can be explained from another perspective. Namely, at high frequencies, where the constructive interference of sound fields of different sources can not be achieved, the least mean squared error solution is biased towards highly attenuating all signals, such that the reconstruction error approaches the desired signal.[5]

A way of avoiding the aforementioned problems—although not providing the correct reproduction in the wide listening area—is normalizing the filters' gains at all frequencies such that on a grid of control points, the average power of the reproduced field is equal to the average power of the desired field. In particular, if $A_1(\omega), \ldots, A_M(\omega)$ and $Y_1(\omega), \ldots, Y_M(\omega)$ are respectively the amplitudes of the desired and the reproduced sound field at frequency ω in M control points, then each loudspeaker filter is corrected by

$$\tilde{H}_i(\omega) = c_f(\omega)H_i(\omega)\,, \tag{10}$$

where $c_f(\omega)$ is a correction factor given by

$$c_f(\omega) = \frac{\sqrt{\sum_{i=1}^{M} A_i^2(\omega)}}{\sqrt{\sum_{i=1}^{M} Y_i^2(\omega)}}\,. \tag{11}$$

2) Loudspeaker subset selection: While it might seem beneficial to use all loudspeakers for reproduction with SFR, there are many cases where using only a subset of loudspeakers can give better reproduction provided the optimization is done for a specific finite listening area. This observation was made for WFS by Verheijen in [13], and was later used by various authors (e.g., Corteel in [16], [17], [23]), where it was shown how

[5]This is a known phenomenon in Wiener filtering, where at low SNRs, the gain of the Wiener filter approaches zero.

based on the location of the primary (reproduced) source and the listening area, one can select a sub-array of loudspeakers which physically contribute the most to the sound field reproduction.

There is a plausible explanation for such a selection. Considering the case where an impulsive sound arrives from the primary source, one expects that at all locations in the listening area, the received sound is of similar duration and consequently without significant spectral impairments. However, using all loudspeakers makes a combination of the impulse responses—due to different delays—more spread in time and more varying across different positions than in the case when only a subset of loudspeakers is used, causing both temporal and spectral deviations.

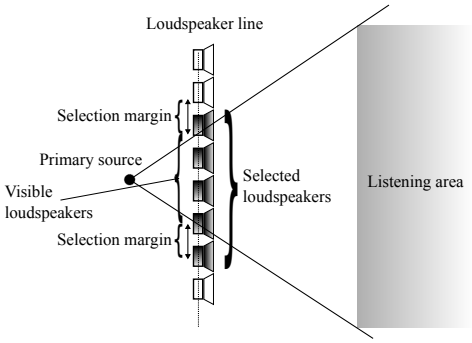


Fig. 3. Illustration of loudspeaker selection based on the primary source position. The visible loudspeakers are inside of the angle subtended by the listening area to the primary source. Loudspeakers within a selection margin of the visible loudspeakers are also selected.

The selection procedure considers only those loudspeakers that are inside—extended by a predefined selection margin—the cone defined by the primary source and the boundaries of the listening area, as shown in Fig. 3. The rationale behind such a choice is twofold: first, it uses the loudspeakers whose contribution is largest when all loudspeakers are used, preserving most of the reconstruction accuracy, and second, the active loudspeakers have lowest delay spreads due to differences in propagation distance and their position relative to the sound wavefront.

Fig. 4 illustrates how reproduction accuracy of SFR does not change notably when only a subset of six (out of 18) loudspeakers is used for reproducing a sinusoid at frequency $f = 500$ Hz. The selected loudspeakers lie in the minimal cone centered at the position of the primary source that contains the listening area. Additionally, loudspeakers outside the minimal cone but within a selection margin can also be selected.

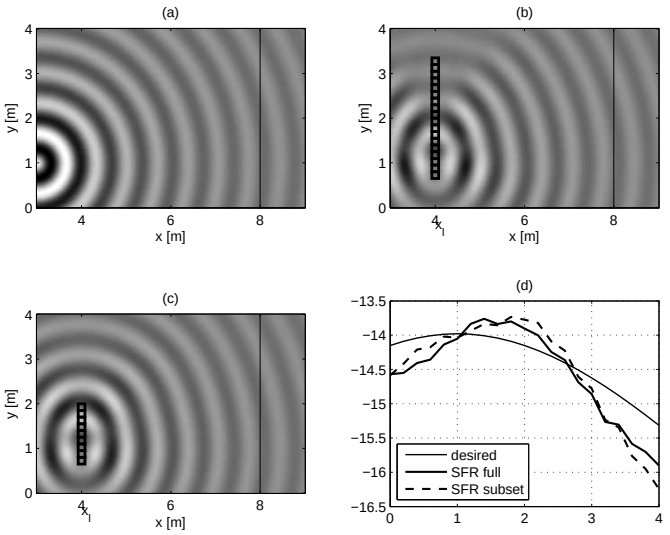


Fig. 4. Comparison of SFR with the entire loudspeaker array from Fig. 9 and SFR with only a sub-array of six selected loudspeakers used to reproduce a point source with frequency $f = 500$ Hz located at $\boldsymbol{r}_m = (3\ \mathrm{m},\ 1\ \mathrm{m})$. The used loudspeakers are marked with squares. (a) Snapshot of the desired sound field; (b) snapshot of the sound field reproduced using all loudspeakers; (c) snapshot of the sound field reproduced with the selected loudspeaker sub-array; (d) magnitude response of the three sound fields on the reference line at frequency $f = 500$ Hz.

3) Control points selection: One also needs to be careful with control point selection, since due to physical limitations set by the loudspeaker array and primary source locations, in some parts of the listening area it is impossible to reproduce the evolution of the desired sound field. Thus, it is physically justified to place control points at locations where the sound wave fronts from the primary source and loudspeakers roughly move towards the same direction.

In the case of primary point sources, control points form a subset of the reference grid which lies inside of a cone defined by the primary source and the active loudspeakers, as shown in Fig. 5(a). For plane wave sources, control points form a subset of the reference grid which lies inside of a stripe defined by the active loudspeakers and the plane wave propagation direction, as shown in Fig. 5(b).

It should also be noted that the selected control points should not lie near the loudspeaker array or the primary source in order to avoid the solution's sensitivity to evanescent (non-propagating) waves. Evanescent waves [29] are a local phenomenon which does not persist with increased distance. Thus, taking sound propagation functions that contain significant evanescent wave energy amounts to model over-fitting and compromises

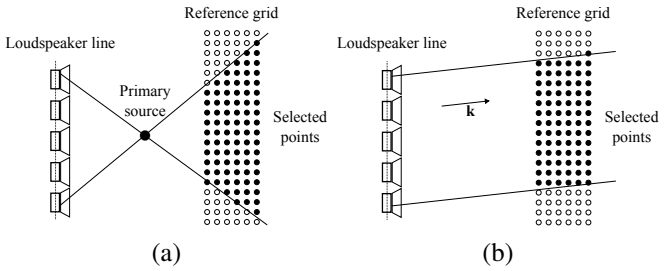


Fig. 5. Illustration of control points selection based on positions of the used loudspeakers and position or direction of the primary source. The selected control points lie inside of the cone or stripe defined by the loudspeaker positions and primary source position or direction, respectively. (a) Point source reproduction; (b) plane source reproduction.

the sound field reproduction accuracy in a larger listening area.

D. Designing discrete-time filters for Sound Field Reconstruction

The frequency-domain SFR filter design procedure uses a non-linear step of discarding small singular values of the system matrix $\boldsymbol{G}(\omega)$. The resulting frequency response and the distribution of singular values at different frequencies are shown in Fig 6. Apparently, SFR filters $\tilde{H}_k(\omega)$ have abrupt changes around frequencies where singular values cross the predefined threshold ϵ. As a consequence, filters $\tilde{H}_k(\omega)$ have long impulse response, which is the main obstacle for designing practical, short discrete-time SFR filters.

However, it turns out that filters $\tilde{h}_k(t)$, being piecewise smooth functions with only a few discontinuities, are well localized in time and most of their energy is concentrated around one main pulse, as shown in Fig. 7. Therefore, shortening $\tilde{h}_k(t)$ does not severely affect the reproduction accuracy, and enables designing efficient discrete-time filters as combinations of a pure delays $\delta_{N_k}[n]$ and a short FIR filters $h_k[n]$ [22].

The SFR discrete-time filter design procedure, illustrated in Fig. 7, use the following three steps:

- *Delay removal:* The main peak of filters $\tilde{h}_k(t)$ can have a long delay for sources far away from the loudspeaker array. In order to avoid using excessively long filters that can accommodate a wide range of different delays, the filters' delays are extracted and realized separately. The delay d_k of the main peak of the SFR filter $\tilde{h}_k(t)$ is extracted considering source-loudspeaker distance and using regression of the phase characteristics of the filter $\tilde{H}_k(\omega)$ [22].

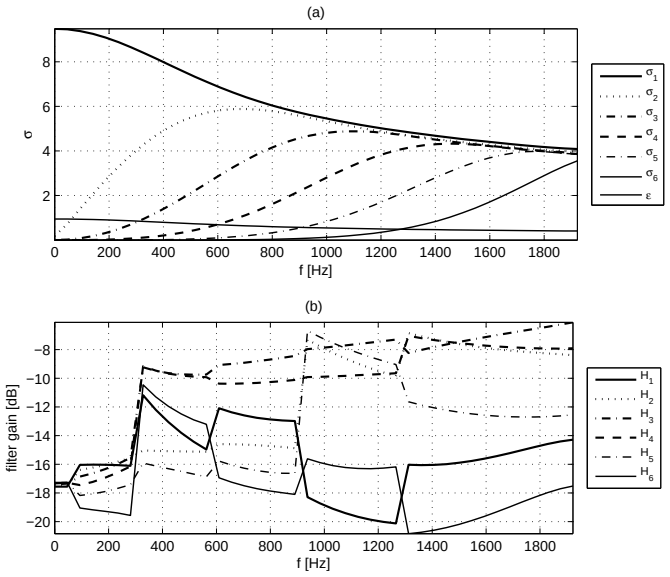


Fig. 6. (a) Singular values of the loudspeaker propagation matrix $\boldsymbol{G}(\omega)$ at different frequencies ω; (b) magnitude response of SFR filters $H_k(\omega)$ obtained from the SFR frequency-domain filter calculation procedure.

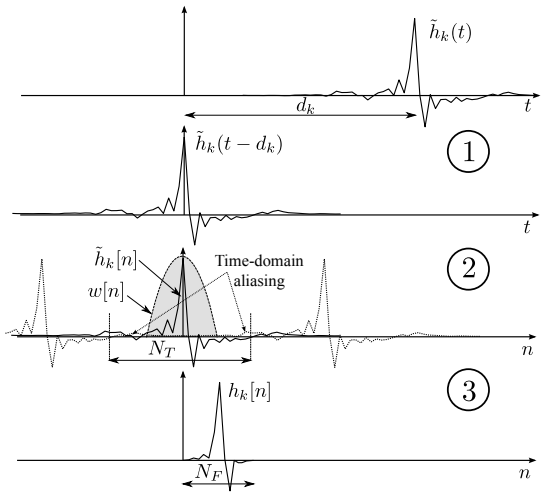


Fig. 7. Conceptual illustration of discrete-time SFR filter design. (1) removing delay d_k; (2) frequency sampling using IDFT; (3) shortening of the filter $\tilde{H}_k(\omega)$, given in (10).

- *Frequency sampling:* At the same time, the problem of frequency sampling of filters $\tilde{H}_k(\omega)$ needs to be solved. In other words, it is necessary to choose the length N_T of the inverse discrete Fourier transform (IDFT) used to obtain the discrete-time impulse response $\tilde{h}_k[n]$. N_T needs to be large enough to give a low time-domain aliasing error. In the setup we used for evaluations, described in the next section, $N_T = 2048$ turned out to be long enough for avoiding notable aliasing artifacts at sampling frequency $f_s = 48$ kHz.

- *Impulse response windowing and delaying:* In the end, $\tilde{h}_k[n]$ is shortened with a tapering window $w[n]$ of length N_F ($N_F < N_T$) and delayed by $N_F/2$ in order to make it causal.

Fig. 8 shows an SFR filter of length $N_F = 512$ samples obtained by the described procedure with IDFT length $N_T = 2048$, for the sampling frequency $f_s = 48$ kHz.

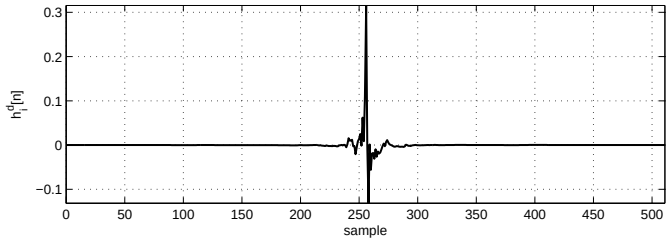


Fig. 8. SFR filter of length $N_F = 512$ samples obtained from frequency response $\tilde{H}_k(\omega)$ using a DFT of length $N_T = 2048$.

III. Evaluation

SFR was evaluated with simulations of the sound reproduction setup shown in Fig. 9. The performance of SFR was compared with two different variants of WFS:

- *WFS I*: Basic WFS, as proposed in the initial works of Berkhout *et al.* [5], [6].
- *WFS II*: WFS that uses loudspeaker selection procedure described in Section II-C2, variants of which were proposed by Verheijen [13] and Corteel [16], [23].

In both variants of WFS, we used a double-sided frequency-independent half-cosine tapering window to mitigate the edge effects. The length of the taper on each end was 15 % of the loudspeaker array length. WFS filters were computed starting from the loudspeaker driving function formulas found in WFS literature [7], [13]. Additionally, a common correction filter $c_f(\omega)$ was computed and applied to all active loudspeakers in order to achieve the desired average power on the points on the reference line. This procedure was described in Section II-C1.

The reproduction setup, which is shown in Fig. 9, consists of 18 loudspeakers spaced at 15 cm. Loudspeakers are modeled as point sources emitting spherical waves. The reproduced primary sources, on the other hand, were modeled as both point and plane sources, emitting spherical and plane waves, respectively.

Two different sets of simulations were performed. The first set gives insight into the reproduction accuracy of the tested approaches both in the frequency and time domain through sound field snapshots. The second set

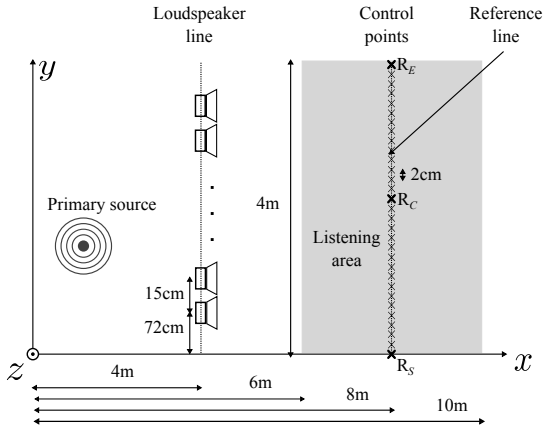


Fig. 9. A sound field reproduction setup using a linear loudspeaker array of 18 loudspeakers spaced at $\Delta_l = 15$ cm. The listening area is a square of size 4 m, 2 m in front of the loudspeaker array.

of simulations gives a more thorough quantitative performance analysis of the tested approaches. It does so by exhaustively analyzing magnitude frequency responses and group delay errors for a large number of reproduced sources and a large number of listening positions.

A. Sound field snapshot analysis

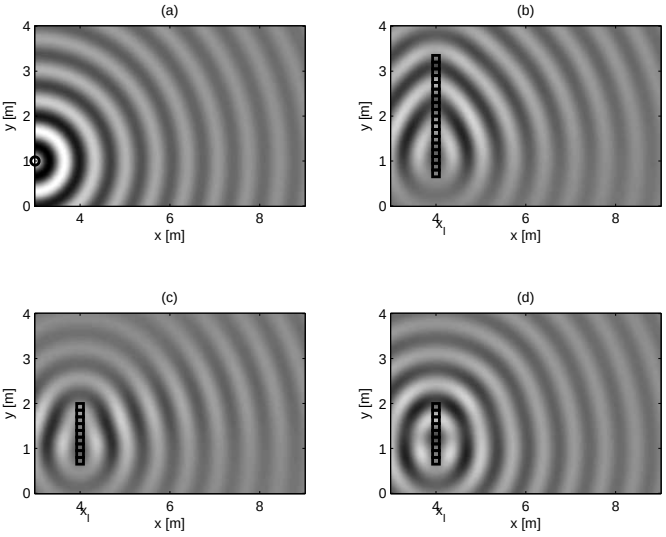


Fig. 10. Comparison of WFS and SFR in reproducing a point source with frequency $f_1 = 500$ Hz located at $\boldsymbol{r}_m = (3\ \mathrm{m},\ 1\ \mathrm{m})$. The used loudspeakers are marked with squares. Sound field snapshots: (a) desired, (b) WFS I, (c) WFS II, and (d) SFR.

1) Sinusoidal sources: The first simulation analyzes the spatial accuracy of reproduction of a sinusoidal (single-frequency) point source. It compares snapshots of the desired and sound fields reproduced with the three

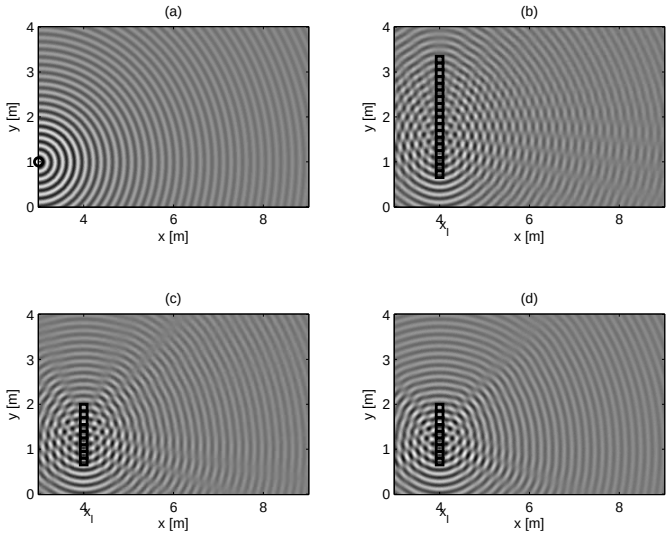


Fig. 11. Comparison of WFS and SFR in reproducing a point source with frequency $f_1 = 2$ kHz located at $\boldsymbol{r}_m = (3\ \mathrm{m},\ 1\ \mathrm{m})$. The used loudspeakers are marked with squares. Sound field snapshots: (a) desired, (b) WFS I, (c) WFS II, and (d) SFR.

tested approaches. Figures 10 and 11 show comparisons for the reproduction of a sinusoidal point source at frequencies $f_1 = 500$ Hz and $f_2 = 2$ kHz, respectively.

Low-frequency reproduction, as can be observed in Fig. 10, is accurate with all three simulated approaches. However, as the frequency increases, aliasing artifacts begin to appear. Fig. 11 shows the difference in the aliasing artifacts between the three approaches. While WFS I has visible aliasing artifacts along the entire listening area (visible as zero responses along multiple directions), SFR and WFS II have only few directional nulls at the periphery of the listening area, and thus preserve spatial reproduction accuracy up to higher frequencies.

2) Low-pass filtered pulse train: The second simulation shows differences between WFS and SFR from a different perspective. Namely, while the first simulation focused on spatial reproduction accuracy as a function of frequency, this simulation focuses on the spatial reproduction accuracy in a wide range of frequencies.

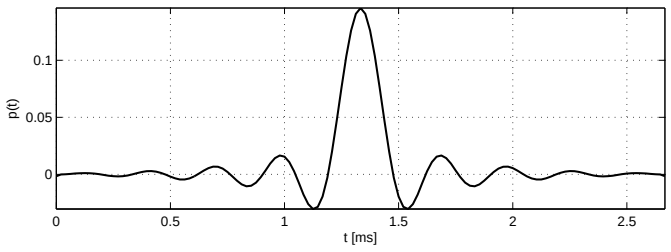


Fig. 12. A low-pass pulse with a cut-off frequency $f_c = 3$ kHz used for constructing a pulse train.

The reproduced primary source is a plane source at angle $\phi = 180^\circ$ emitting a train of low-pass filtered pulses $p(t)$ spaced in time by $T_p = 4$ ms. The shape of a single pulse is shown in Fig. 12.

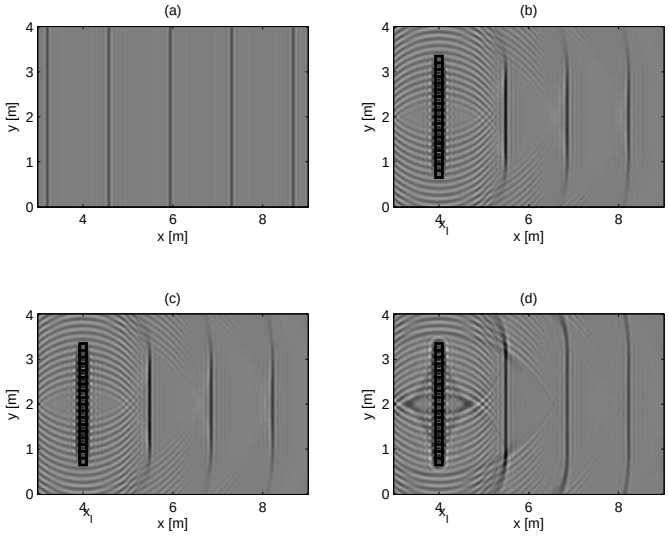


Fig. 13. Comparison of WFS and SFR in reproducing a plane source located at an angle $\phi = \pi$ that emits a train of low-pass pulses with a period of $T_p = 4$ ms. The used loudspeakers are marked with squares. Sound field snapshots: (a) desired, (b) WFS I, (c) WFS II, and (d) SFR.

Fig. 13 shows snapshots of the desired sound field and the sound fields reproduced with SFR and the two variants of WFS. Note that in this scenario loudspeaker selection has no effect and as a consequence all three approaches use all loudspeakers and the two variants of WFS are identical. From the snapshots of the reproduced fields, it is apparent that the shape of sound wave fronts is accurately reproduced across the listening area with both WFS and SFR. However, observing the amplitude of the emitted pulses across the listening area, one can see that with WFS amplitude notably decreases towards the sides. SFR, on the other hand, does not suffer from this problem.

Fig. 14, which shows magnitude frequency responses in the center and at both ends of the listening line (located four meters in front of the loudspeaker array), corroborates the previous visual observation from the sound field snapshots. In particular, it shows how SFR's low-frequency magnitude response at both ends of the listening line is flatter and less attenuated relative to the desired characteristics when compared to WFS.

At this point, some preliminary conclusions can be drawn about the advantages of SFR over WFS. Compared to WFS I, SFR provides more graceful degradation of reproduction accuracy across the listening area as the frequency increases. This effectively means that

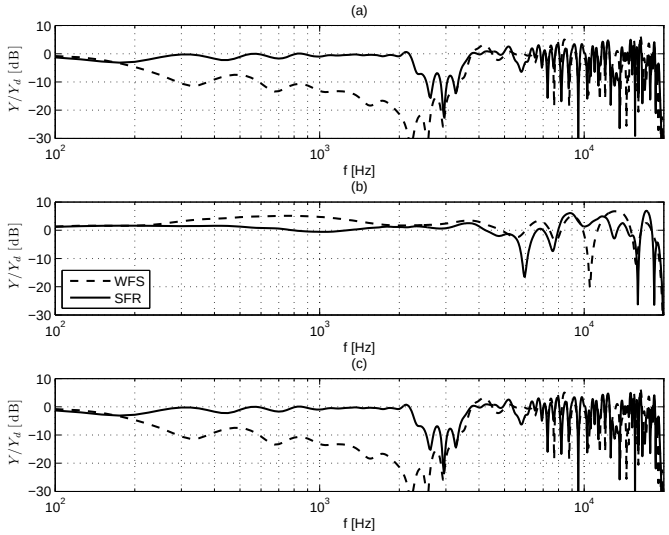


Fig. 14. Normalized magnitude frequency responses of SFR and WFS in the control points (a) $R_S(8\ m,\ 0)$, (b) $R_C(8\ m,\ 2\ m)$, and (c) $R_E(8\ m,\ 4\ m)$ relative to a plane source located at an angle $\phi = \pi$.

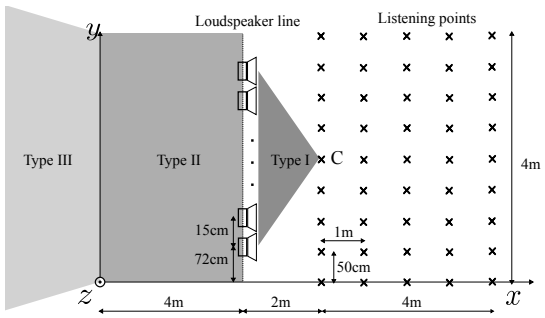


Fig. 15. Sound field reproduction setup showing three categories of simulated primary sources: Type I, Type II, and Type III, and a grid of listening points covering the listening area, where responses are computed.

compared to WFS I, SFR increases both the aliasing frequency margin and enlarges the effective listening area. Furthermore, loudspeaker subset selection in WFS II helps decreasing the aliasing artifacts as the frequency increases, but as will be shown next, this improvement comes at the cost of increasing average magnitude spectral deviations across the listening space.

B. Impulse response analysis

In order to remove the influence of particular source and listening positions and make a more general observation about the performance of WFS and SFR, we performed a number of simulations involving multiple primary source and listening positions. The simulated primary sources—30 altogether—were divided into three different categories, with each category containing ten sources (see Fig. 15):

- *Type I*: Focused, frontal point sources located inside of a triangle whose vertexes coincide with two outer loudspeakers and the point $C(6\ \mathrm{m},\ 2\ \mathrm{m})$, regularly spaced along the x axis and with y coordinates chosen uniformly at random within the triangle boundaries.
- *Type II*: Point sources located closely behind the loudspeaker array. In the simulations, these sources were regularly spaced along the x axis between $x = 0$ and $x = 4$ m, and their y coordinates were chosen uniformly at random between $y = 0$ and $y = 4$ m.
- *Type III*: Point sources far away from the loudspeaker array, which were modeled as sources emitting plane waves. In the performed simulations, these plane wave sources were positioned at ten regularly-spaced directions between $165°$ and $180°$.

The three categories of simulated primary sources are illustrated in Fig. 15

The impulse responses for each simulated primary source were computed on a finite rectangular grid of listening points spaced at 1 m along the x axis and 50 cm along the y axis, shown in Fig. 15.

For each primary source category, we formed aggregated plots containing statistics of normalized magnitude frequency responses and group delay errors for all listening points and all primary sources in the category.

The normalized magnitude frequency responses are given by

$$Y_n(f) = \frac{Y(f)}{Y_d(f)}\,, \tag{12}$$

where $Y(f)$ is the reproduced field's magnitude response in a listening point and $Y_d(f)$ is the desired magnitude response in that point.

The group delay error $e_\tau(f)$ is given by the difference between the group delay $\tau_g(f)$ of the reproduced impulse response in a listening point and the group delay $\tau_g^d(f)$ of the desired impulse response in that point

$$e_\tau(f) = \tau_g(f) - \tau_g^d(f)\,. \tag{13}$$

The plots contain $5 - 95$ percentiles, $25 - 75$ percentiles, and the median value of the said quantities across the audible frequency range.

The aggregated statistical plots of magnitude and group delays provide insight not only into how accurate the tested approaches are on average, but they also show to what extent reproduction accuracy varies across space.

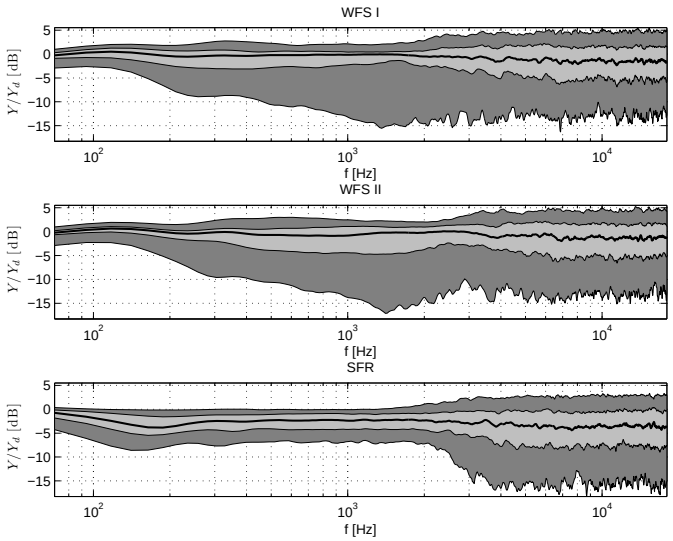


Fig. 16. Normalized magnitude frequency responses of WFS I, WFS II, and SFR for focused point sources (Type I) on a grid of listening points. Light-gray area shows $25-75$ percentiles, dark-gray area shows $5-95$ percentiles, and solid black line shows the median.

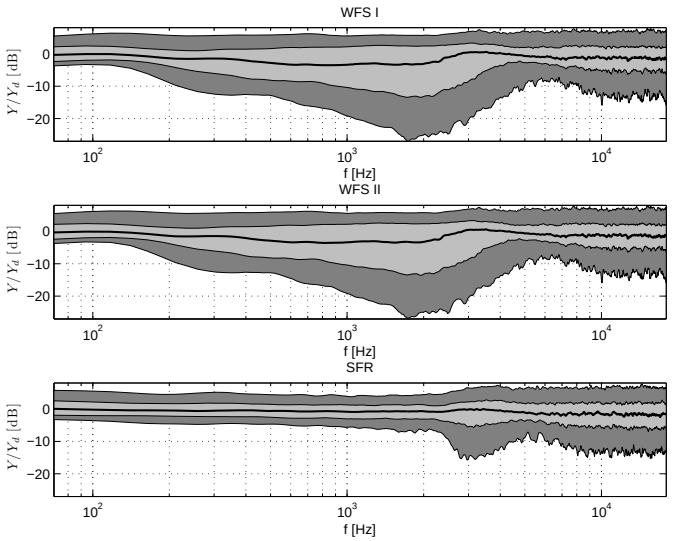


Fig. 18. Normalized magnitude frequency responses of WFS I, WFS II, and SFR for plane sources at directions $\alpha \in [165°, 180°]$ (Type III) on a grid of listening points. Light-gray area shows $25-75$ percentiles, dark-gray area shows $5-95$ percentiles, and solid black line shows the median.

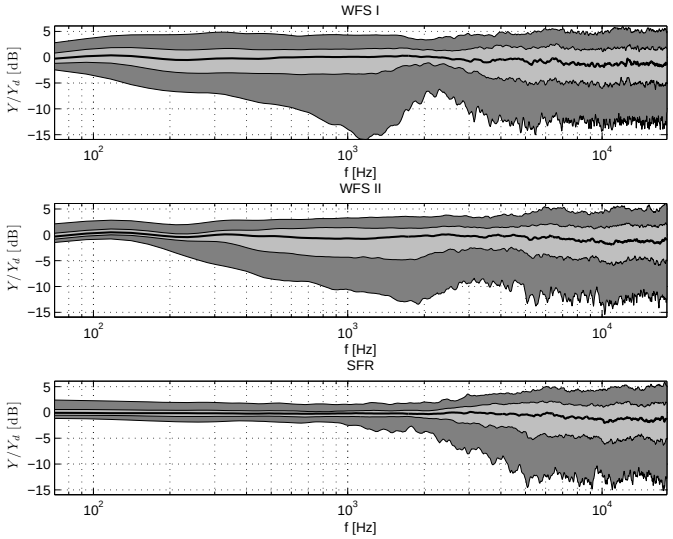


Fig. 17. Normalized magnitude frequency responses of WFS I, WFS II, and SFR for point sources behind the loudspeaker array (Type II) on a grid of listening points. Light-gray area shows $25-75$ percentiles, dark-gray area shows $5-95$ percentiles, and solid black line shows the median.

1) Statistical magnitude frequency response plots: Figures 16, 17, and 18 show the previously described magnitude frequency response statistical plots for primary sources of Type I, II, and III, respectively. It can be seen that with SFR, the $25-75$ percentiles of magnitude frequency responses are within 2 dB of the desired responses up to around 4 kHz for all three primary source categories. The median of SFR's normalized magnitude response lies at 0 dB across low frequencies, as opposed to the median magnitude responses of the two WFS approaches, which vary around 0 dB. Although $5-95$ percentiles exhibit variations around the median up to around 10 dB, meaning that for some source-listening position pairs the reproduced impulse response differs significantly from the desired one, they are notably smaller than the variations of the corresponding percentiles of magnitude frequency responses of the two WFS approaches. Above the frequency of 4 kHz, the three approaches perform similarly due to spatial aliasing.

It should be noted that WFS II exhibits more spectral magnitude artifacts in the extended listening area compared to WFS I. It can thus be said that the previously observed improvement of aliasing performance, apparent in Fig. 11, comes at the cost of reducing the listening area size. This observation was also reported by Corteel *et al.* in [23].

2) Statistical group delay error plots: Figures 19, 20, and 21 show the group delay error statistical plots for primary sources of Type I, II, and III, respectively.

From Figures 19 and 21, it can be observed that the two WFS approaches have virtually the same group delay performance for focused (Type I) and plane (Type III) sources. This is not surprising, as for most of the simulated focused and plane sources, both WFS approaches use all loudspeakers. For focused and plane sources, SFR's group delay performance is on average better or

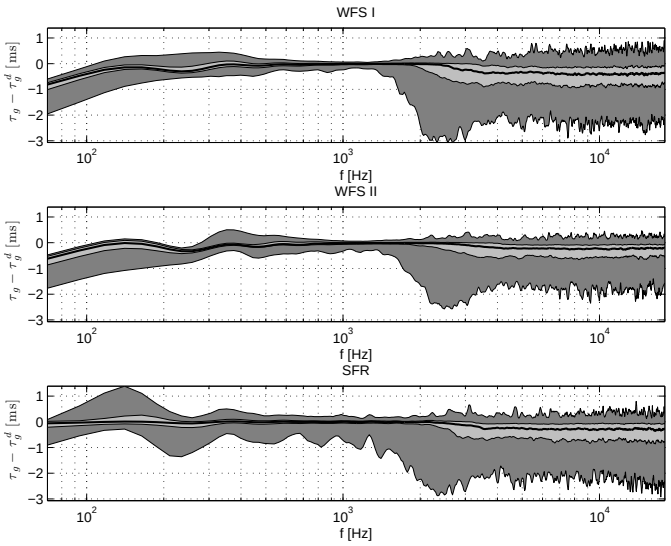


Fig. 19. Group delay errors $e_\tau(f) = \tau_g(f) - \tau_g^d(f)$ of WFS I, WFS II, and SFR for focused point sources (Type I) on a grid of listening points. Light-gray area shows $25-75$ percentiles, dark-gray area shows $5-95$ percentiles, and solid black line shows the median.

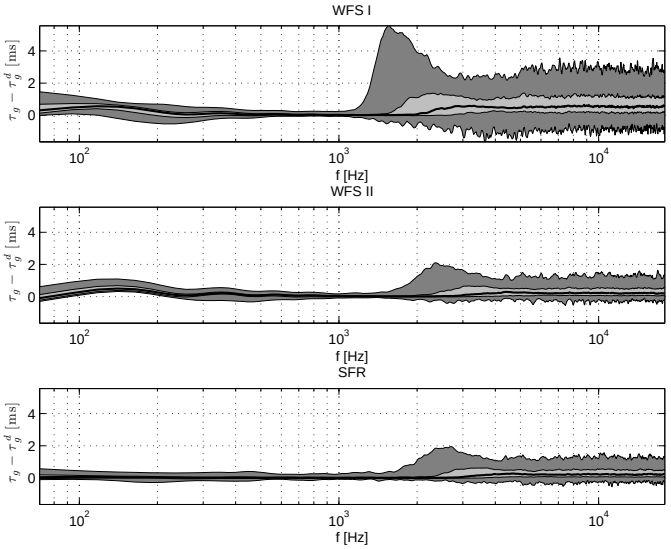


Fig. 20. Group delay errors $e_\tau(f) = \tau_g(f) - \tau_g^d(f)$ of WFS I, WFS II, and SFR for point sources behind the loudspeaker array (Type II) on a grid of listening points. Light-gray area shows $25-75$ percentiles, dark-gray area shows $5-95$ percentiles, and solid black line shows the median.

comparable to both WFS approaches, as can be observed from $25-75$ group delay error percentiles. SFR's group delay error is more variable in the extreme cases in the frequency range $500-2000$ Hz, which is apparent from observing $5-95$ percentiles. Note, however, that the group delay errors in the frequency range $500-2000$ Hz are below the group delay discrimination threshold of

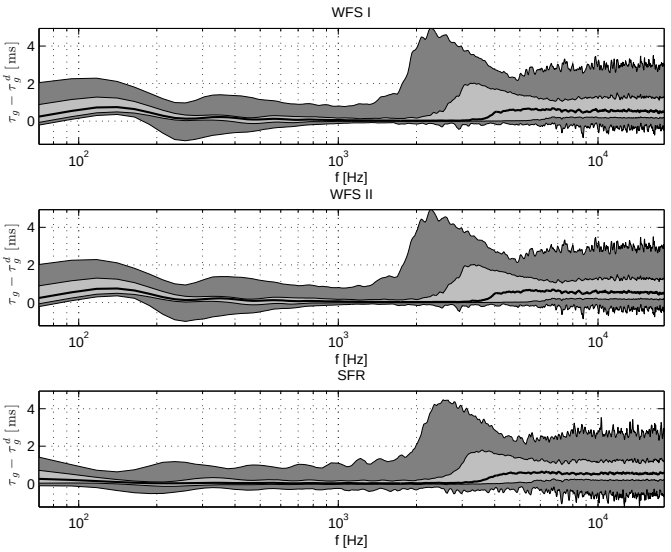


Fig. 21. Group delay errors $e_\tau(f) = \tau_g(f) - \tau_g^d(f)$ of WFS I, WFS II, and SFR for plane sources at directions $\alpha \in [165°, 180°]$ (Type III) on a grid of listening points. Light-gray area shows $25-75$ percentiles, dark-gray area shows $5-95$ percentiles, and solid black line shows the median.

2 ms, as found by Flanagan *et al.* [30] [6] Therefore, a slightly higher group delay variance of SFR should not be a cause of notable perceptual artifacts.

Fig. 20 shows that the group delay performance of WFS II is superior to WFS I when reproducing point sources behind the loudspeaker array (Type II sources). SFR, on the other hand, has group delay errors which are similar to WFS II: on average, SFR is slightly better, but also slightly more variable in the frequency range $500-2000$ Hz.

C. Discussion

From the frequency-domain analysis of the impulse responses on a grid of listening points, three observation can be made.

- It is apparent that the low-frequency response of SFR exhibits little spectral deviations up to almost 4 kHz for all categories of simulated primary sources and all listening points. Both WFS variants, on the other hand, suffer from higher coloration artifacts across space in the low-frequency range.
- It can be seen that more notable spectral deviations across space start at higher frequencies with SFR when compared to both WFS approaches.
- The average group delay performance of SFR is slightly better than the two WFS variants, but slightly more variable in the low-frequency range

[6] See also [31].

$500 - 2000$ Hz. Nevertheless, the range of group delay error variations in this frequency range is below the group delay discrimination threshold of 2 ms [30].

The presented extensive comparisons confirm the previous observation that SFR provides an effective extension of the listening area with correct sound field reproduction. Additionally, it raises the reproduction aliasing frequency when compared with both variants of WFS, as described in Section II-C2.

IV. Practical considerations

A. Computational complexity

The presented approach to reproducing physical sound fields has an appealing performance in terms of reproduction accuracy. However, it comes at a cost of increased complexity that stems from solving an MIMO inversion problem in the frequency domain. For each virtual source, the reproduction system needs to perform SVD of the $M \times L$ matrix $\boldsymbol{G}(\omega)$ at $\frac{N_F}{2}+1$ frequencies.[7] Since the number of control points M is usually larger than the number of loudspeakers L, the complexity of obtaining SFR filters for one virtual source is given by $\Theta(M^2 L N_F)$.

High computational complexity makes real-time calculation of loudspeaker filters with SFR, like with most numerical approaches, difficult. Instead, one can produce a database of reproduction filters offline, which can then be read in real time for sound field rendering purposes. This would entail dividing the reproduction zone with a polygonal mesh and pre-computing the filters corresponding to every element of the mesh. In the simplest case of a uniform rectangular mesh, the SFR filters for a single virtual source can be obtained in constant time based on its position. If the rectangular mesh is non-uniform, the filters are obtained in the time it takes to locate the rectangle that contains the virtual source. This complexity is $\Theta(\log N_x)$, where N_x is the number of rectangles along the dimension that contains more rectangle "stripes". For a general mesh, the use of space partitioning data structures, such as kd-trees, makes the complexity of obtaining SFR filters $\Theta(\log N_M)$, where N_M is the mesh size.

The previously mentioned filter pre-computation methods allow for real-time sound field rendering, and have already been proposed and used in practical multichannel sound field reproduction systems.

[7]Real-valued filters have conjugate-symmetric spectra.

B. Performing system measurements

It has already been stressed that SFR does not put limiting constraints on the reproduction setup. It works irrespectively of loudspeaker or desired source directivity, loudspeaker calibration, or sound propagation characteristics, as long as one is able to obtain the MIMO acoustic channel involving a dense grid of control points. In a practical reproduction system, this requirement might be too strict, as it is hard to imagine that one would measure loudspeaker responses on a fine grid, especially in larger venues.

Instead, one could compromise by at least measuring the system on a contour in the reproduction plane that encloses the listening area. By doing so, one trades off some reproduction accuracy for practicality. Simulation experiments involving enclosing contours instead of covering grids, not presented in this paper for the sake of space, show that SFR does not suffer from a noticeable performance loss with the said simplification.

V. Conclusions

This paper described Sound Field Reconstruction, a technique for reproducing sound fields in an extended listening area using an array of loudspeakers. SFR is based on a numerical optimization procedure for MIMO channel inversion. The control points covering the listening area, used by the MIMO channel inversion procedure, are spaced below the Nyquist criterion to avoid aliasing. Additionally, SFR uses geometry-based loudspeaker and control points selection to mitigate artifacts due to aliasing and over-fitting.

SFR is a flexible sound field reproduction approach applicable to loudspeaker arrays with different topologies and directivities. It also enables reproducing directive sources, and does not restrict applications to free-field or anechoic sound propagation conditions.

We have shown that, compared to Wave Field Synthesis, which is the state of the art technique for sound field reproduction using loudspeaker arrays, SFR achieves better average reproduction accuracy in an extended listening area and preserves reproduction accuracy up to higher frequencies.

VI. Acknowledgment

This work was supported by the National Competence Center in Research on Mobile Information and Communication Systems (NCCR-MICS, http://wwww.mics.org), a center supported by the Swiss National Science Foundation under grant number 5005-67322. It was also supported by the Swiss National Science Foundation grant "New Sampling Methods for Processing and Communication", grant number 200021-121935.

REFERENCES

[1] A. Blumlein, "Improvements in and relating to sound transmission, sound recording and sound reproduction systems," *British Patent Specification 394325*, 1931, Reprinted in *Stereophonic Techniques*, Aud. Eng. Soc, New York, 1986.

[2] P.S. Gaskel and P.A. Ratliff, "Quadraphony: developments in matrix H decoding," Tech. Rep., BBC Research Department, Feb 1977.

[3] M.A. Gerzon, "Periphony: width-height sound reproduction," *J. Aud. Eng. Soc*, vol. 21, no. 1, pp. 2–10, Jan 1973.

[4] M.A. Gerzon, "Practical Periphony: the reproduction of full-sphere sound," in *Preprints 65th Conv. Aud. Eng. Soc*, Feb 1980.

[5] A.J. Berkhout, "A holographic approach to acoustic control," *J. Audio Eng. Soc*, vol. 36, no. 12, pp. 977–995, 1988.

[6] A.J. Berkhout, D. de Vries, and P. Vogel, "Wave front synthesis: a new direction in electroacoustics," in *Preprints 93th Conv. Aud. Eng. Soc*, Oct 1992.

[7] A.J. Berkhout, D. de Vries, and P. Vogel, "Acoustic control by Wave Field Synthesis," *J. Acoust. Soc. Am*, vol. 93, pp. 2764–2764, 1993.

[8] D.H. Cooper and T. Shiga, "Discrete-matrix multichannel stereo," *J. Audio Eng. Soc*, vol. 20, no. 5, pp. 346–360, 1972.

[9] J.S. Bamford and J. Vanderkooy, "Ambisonic sound for us," in *Preprints 99th Conv. Aud. Eng. Soc*, 1995.

[10] J. Daniel, R. Nicol, and S. Moreau, "Further investigations of high order Ambisonics and wavefield synthesis for holophonic sound imaging," in *Preprints 114th Conv. Aud. Eng. Soc*, 2003.

[11] J. Blauert, *Spatial hearing: the psychophysics of human sound localization*, MIT Press, 1997.

[12] E.W. Start, "Application of curved arrays in Wave Field Synthesis," in *Preprints 100th Conv. Aud. Eng. Soc*, 1996.

[13] E.N.G. Verheijen, *Sound reproduction by Wave Field Synthesis*, Ph.D. thesis, Delft University of Technology, 1997.

[14] D. de Vries, "Sound reinforcement by wavefield synthesis: adaptation of the synthesis operator to the loudspeaker directivity characteristics," *J. Audio Eng. Soc*, vol. 44, no. 12, pp. 1120–1131, 1996.

[15] J. Ahrens and S. Spors, "Sound field reproduction using planar and linear arrays of loudspeakers," *IEEE Trans. Audio Speech Lang. Proc*, vol. 18, no. 8, pp. 2038–2050, Nov 2010.

[16] E. Corteel, "Equalization in an extended area using multichannel inversion and Wave Field Synthesis," *J. Audio Eng. Soc*, vol. 54, no. 12, pp. 1140–1161, 2006.

[17] E. Corteel, "Synthesis of directional sources using Wave Field Synthesis, possibilities, and limitations," *EURASIP J. App. Sig. Proc*, vol. 2007, no. 1, pp. 188–188, 2007.

[18] S. Spors, "Extension of an analytic secondary source selection criterion for Wave Field Synthesis," in *Preprints 123th Conv. Aud. Eng. Soc*, 2007.

[19] O. Kirkeby and P.A. Nelson, "Reproduction of plane wave sound fields," *J. Acoust. Soc. Am*, vol. 94, pp. 2992–3000, 1993.

[20] P.A. Gauthier and A. Berry, "Adaptive wave field synthesis for sound field reproduction: Theory, experiments, and future perspectives," *J. Audio Eng. Soc*, vol. 55, no. 12, pp. 1107–1124, 2007.

[21] M. Kolundžija, C. Faller, and M. Vetterli, "Sound Field Reconstruction: an improved approach for wave field synthesis," in *Preprints 126th Conv. Aud. Eng. Soc.*, May 2009.

[22] M. Kolundžija, C. Faller, and M. Vetterli, "Designing practical filters for Sound Field Reconstruction," in *Preprints 127th Conv. Aud. Eng. Soc*, Oct 2009.

[23] E. Corteel, R. Pellegrini, and C. Kuhn-Rahloff, "Wave Field Synthesis with increased aliasing frequency," in *Preprints 124th Conv. Aud. Eng. Soc*, May 2008.

[24] T. Ajdler, L. Sbaiz, and M. Vetterli, "The plenacoustic function and its sampling," *IEEE Trans. Sig. Proc*, vol. 54, no. 10, pp. 3790–3804, 2006.

[25] O. Kirkeby, P.A. Nelson, H. Hamada, and F. Orduna-Bustamante, "Fast deconvolution of multichannel systems using regularization," *IEEE Trans. Speech Audio Proc*, vol. 6, no. 2, pp. 189–194, Mar 1998.

[26] P.A. Nelson and S.J. Elliott, *Active control of sound*, Academic Press, 1992.

[27] J. Hannemann and K.D. Donohue, "Virtual sound source rendering using a multipole-expansion and method-of-moments approach," *J. Audio Eng. Soc*, vol. 56, no. 6, pp. 473–481, 2008.

[28] G. Golub and W. Kahan, "Calculating the singular values and pseudo-inverse of a matrix," *J. Soc. Ind. Appl. Math*, vol. 2, no. 2, pp. 205–224, 1965.

[29] E.G. Williams, *Fourier acoustics: sound radiation and nearfield acoustical holography*, Academic Press, 1999.

[30] S. Flanagan, B.C.J. Moore, and M.A. Stone, "Discrimination of group delay in clicklike signals presented via headphones and loudspeakers," *J. Audio Eng. Soc*, vol. 53, no. 7-8, pp. 593–611, 2005.

[31] J. Blauert and P. Laws, "Group delay distortions in electroacoustical systems," *J. Acoust. Soc. Am*, vol. 63, pp. 1478–1483, 1978.